

# Probabilistic Rateless Multiple Access for Machine-to-Machine Communication

Mahyar Shirvanimoghaddam, *Member, IEEE*, Yonghui Li, *Senior Member, IEEE*, Mischa Dohler, *Fellow, IEEE*, and Branka Vucetic, *Fellow, IEEE*,

**Abstract**—Future machine to machine (M2M) communications need to support a massive number of devices communicating with each other with little or no human intervention. Random access techniques were originally proposed to enable M2M multiple access, but suffer from severe congestion and access delay in an M2M system with a large number of devices. In this paper, we propose a novel multiple access scheme for M2M communications based on the capacity-approaching analog fountain code to efficiently minimize the access delay and satisfy the delay requirement for each device. This is achieved by allowing M2M devices to transmit at the same time on the same channel in an optimal probabilistic manner based on their individual delay requirements. Simulation results show that the proposed scheme achieves a near optimal rate performance and at the same time guarantees the delay requirements of the devices. We further propose a simple random access strategy and characterized the required overhead. Simulation results show the proposed approach significantly outperforms the existing random access schemes currently used in long term evolution advanced (LTE-A) standard in terms of the access delay.

**Index Terms**—Analog fountain codes, machine-to-machine, message passing decoder, massive multiple access.

## I. INTRODUCTION

MACHINE-to-Machine communications have emerged as a promising technology to enable trillions of multi-role devices, namely machine-type communications (MTC) devices, to communicate with each other with little or no human intervention [1, 2]. It has many potential applications, such as intelligent transportation systems (ITS), healthcare monitoring, retail, banking, smart grids, home automation and so on. It is expected that in the next a few years over 2 billion MTC devices will become directly attached to cellular networks to provide M2M communications [3]. Thus, there will be a massive number of MTC devices with no/low mobility [4] in each cell, which is significantly more than the number of users in current cellular networks. Moreover, M2M traffic involves a large number of short-lived sessions, attempting to deliver a small amount of data (few hundred bits) to the base station, which is quite different from those in human-to-human (H2H) communications. Such differences motivate researchers around the globe to optimize the current cellular networks to effectively enable M2M communications [5, 6].

The material in this paper was presented in part at the 2014 IEEE International Symposium on Information Theory, Honolulu, HI.

M. Shirvanimoghaddam, Y. Li, and B. Vucetic are with the Center of Excellence in Telecommunications, School of Electrical and Information Engineering, The University of Sydney, Sydney, NSW 2006, Australia (e-mail: mahyar.shirvanimoghaddam@sydney.edu.au; yonghui.li@sydney.edu.au; branka.vucetic@sydney.edu.au).

M. Dohler is with King's College London (e-mail: mischa.dohler@kcl.ac.uk)

## A. Motivation and Related Work

In most existing wireless access networks, the first step in establishing an air interface connection, is to perform random access (RA) in a contention manner. In fact, short-lived sessions with small amount of data in M2M communications, makes it inefficient to establish dedicated resources for data transmission [7]. Such a random access approach will not work effectively when the number of nodes is very large, due to frequent transmission collisions, leading to network congestion, unexpected delays, packet loss, high energy consumption, and radio resource wastage [7]. Thus, one of critical challenges in M2M communications is to design an effective medium access scheme to support a large number of devices. In current RA approaches, when two or more devices select the same RA preamble in the first phase of the contention phase, a collision will occur and the respective devices will not be scheduled for data transmission. To reduce the access delay in M2M communications, several schemes have been proposed, such as *dynamic allocation* [8, 9], *slotted access*, *group-based* [1, 10], *pull-based*, and *access class barring* techniques [11, 12]. Although these approaches can reduce the access collisions to a certain degree, most of them still suffer from very high access delays in highly dense networks. The main idea behind these schemes is to delay the retransmission of the access request for a random/fixed amount of time, thus increasing the access probability within a relatively short time. This is however inefficient in M2M communications due to small short burst transmissions of devices which mainly do not require the whole RB for their transmission. This means that for M2M communications with a very large number of devices there might not be enough RBs to be orthogonally allocated to the devices, which significantly increase the access delay even if the random access requests are delivered correctly.

Additionally, different MTC devices have diverse service requirements and traffic patterns in M2M communications. Generally, we can divide the traffic types into four different categories. The first type is the alarm traffic, which is completely random and its probability is very low; however, it has a very strict delay requirement. The second traffic type can be modeled by a random Poisson distribution with the parameters depending on the application [13]. The regular traffic, such as smart metering applications, is the third traffic type, and the last type is the streaming, like video surveillance applications. Current proposals for enabling M2M communications, did not consider the priorities among devices and different quality of service (QoS) requirements. These approaches are mostly inefficient for M2M communications as they are generally designed for a fixed payload size and thus, cannot support

M2M applications with different service requirements.

Recently, a systematic framework has been developed in [13, 14] to understand the fundamental limits of M2M communications in terms of power efficiency and throughput. However, they did not provide a systematic approach to develop an efficient communication protocol to approach these limits. Here, we consider a realistic model for M2M communications, which supports both regular and random traffics with different delay and service requirements. We develop a practical transmission scheme for M2M communications based on recently proposed analog fountain codes (AFCs) [15] to enable massive number of devices communicating with a common base station (BS) while satisfying QoS requirements of all devices. We further show that the proposed scheme can closely approach the fundamental limits of M2M communications in terms of throughput and can satisfy the delay requirements of all devices at the same time. The main contributions of this paper are summarized next.

### B. Contributions

1) *Efficient Random Access Proposal*: We consider the contention-based random access in the proposed probabilistic multiple access for M2M communications. In the contention process, each device transmits an RA preamble to the BS to establish a communication link with the BS. In existing RA schemes in M2M communications, the devices will be identified by the BS in the contention phase if each device has chosen an RA preamble which is not selected by other devices, and a resource block (RB) will be allocated to the device for the data transmission. For a system with a large number of devices, several devices may choose the same RA preamble and thus the BS cannot recognize the devices in the RA process. Moreover, even if the BS can identify all active devices, the number of RBs is not sufficient to support all of them. To overcome this problem, we propose a new contention mechanism. We group the RA preambles into several subsets according to delay requirements of MTC devices. Then, a sub-set of RA preambles is assigned for those devices which have the same delay requirement. In the contention period, each device selects a specific RA preamble based on its delay requirement from the associated subset of RA preambles. By detecting the RA preamble of the device, the BS then knows which subset this RA preamble belongs to and thus knows its delay requirement. The BS detects the number of devices which have selected the same RA preamble and broadcasts this information to the devices. The devices which have selected the same RA preamble then transmit in the same RB, and their identities will be transmitted along with their payload data and will be later recognized by the BS after the decoding. In this way, each available RB will be effectively used by a number of devices and RA collisions is handled in an efficient manner to simultaneously support collided devices.

2) *Efficient Maximum Throughput Transmission*: Our second contribution in this work is to represent the multiple access process as a special kind of analog graph code. This enables us to draw on the recently proposed capacity approaching analog fountain codes [15] to design the optimal access protocol achieving the sum-rate capacity of multiple access channel.

Fountain codes have been used to facilitate the multiple access transmissions over erasure channels [16, 17, 18, 19], where collided packets are used in an effective way to increase the overall system throughput; however, they cannot be used for wireless channels which are affected by fading and noise. In [20], it has been shown that the capacity of a Gaussian multiple access channel can be achieved by using spatially coupled sparse graph multi-user modulation with an iterative interference cancelation scheme at the receiver. However, this approach performs decoding and interference cancelation separately and its complexity increases significantly as the number of access users increases. Thus, it will not be feasible for M2M applications with a massive number of users. To overcome this problem, we proposed a simple capacity-approaching analog fountain code (AFC) in [15] with linear complexity in both the encoder and decoder. Thanks to the interesting linear property of AFCs, the summation of AFC coded signals of multiple simultaneous transmitted users still forms an AFC code [21]. Thus the receiver can represent the multiple access transmission by a special kind of graph codes, and the BS can jointly decode them by using a standard belief propagation (BP) decoder. The proposed probabilistic multiple access AFC (MA-AFC) scheme enables the devices to transmit in a probabilistic manner in a way to form an equivalent AFC code at the BS, and approach the sum-rate capacity of the multiple access channel [21].

3) *QoS aware Random Transmission*: Different devices in M2M communications have diverse traffic and service requirements. Existing random access schemes do not provide QoS guarantees in the random access transmissions. This precludes use in many practical M2M systems, where timing constraints are critical. How to alleviate access congestions in massive access transmissions while providing delay or QoS guarantees remains a significant hurdle in M2M access network design. In the proposed approach, the degree distribution functions of AFCs and access probabilities are optimized such that the delay requirement for each device is satisfied. We formulate an optimization problem based on the density evolution analysis of the AFC code to find the optimum degree and access probability for each device to satisfy their delay requirements. We further show that the proposed approach can simultaneously satisfy the delay requirement for all devices.

Throughout the paper the following notations are used. We use a boldface small letter to denote a vector and the  $i^{th}$  entry of vector  $\mathbf{v}$  is denoted by  $v_i$ . A boldface capital letter is used to denote a matrix and the  $i^{th}$  element in the  $j^{th}$  row of matrix  $\mathbf{G}$  is denoted by  $g_{i,j}$ . The transpose of matrix  $\mathbf{G}$  is denoted by  $\mathbf{G}'$ . A set is denoted by a curly capital letter and the cardinality of set  $\mathcal{S}$  is represented by  $|\mathcal{S}|$ .

The rest of the paper is organized as follows. In Section II, we present the system model. The proposed probabilistic multiple access scheme for M2M communications is presented in Section III. Section IV presents the asymptotic analysis of the belief propagation decoding of the MA-AFC scheme based on the density evolution analysis. In Section V, an optimization problem is formulated to find the optimum code parameters. Simulation results are shown in Section VI, and finally Section VII concludes the paper.

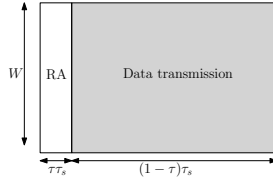


Fig. 1. A resource block (RB) of duration  $\tau_s$  and bandwidth  $W$ .

## II. SYSTEM MODEL

In this section, we first introduce the network and channel model in M2M communications. Then we discuss in more details the packet transmission and delay model for MTC devices.

### A. Network Model and Random Access

We consider an M2M last mile access system, where  $N$  MTC devices are uniformly distributed in a single cell of radius  $r_0$  and communicate with a base station (BS) located at the origin. In this paper, we focus on the M2M applications where MTC devices are deployed at fixed locations, such as smart meters, sensors or cameras in buildings, roads, bridges, etc. [22].

It is assumed that the channel between each device and the BS is a slow time-varying block fading channel, for which the channel remains constant within one transmission block but varies slowly from one block to the other. We consider a time division duplex (TDD)-based wireless access system, where the channel gain of the uplink is assumed to be the same as that of the downlink [23]. With this assumption, each device can estimate the uplink channel gain from the pilot signal sent periodically over the downlink channel by the BS. Using the pilot signal, MTC devices also synchronize their timing to that of the BS. The BS, however, does not have the knowledge of any channel state information (CSI). This assumption is particularly relevant in M2M communications, where due to a large number of devices, it would be impractical for the BS to obtain CSI to every MTC device [24]. The channel between each device and the BS is modeled by path loss, large scale shadowing and small scale fading effects. Thus, the received power at the BS for the signal transmitted from  $U_i$  is given by:

$$P_{r,i} = P_{t,i} \gamma |h_i|^2, \quad (1)$$

where  $P_{t,i}$  is the transmit power,  $\gamma$  is the reference signal to noise ratio (SNR), defined as the average received SNR from a device transmitting at maximum power located at the cell edge,  $h_i$  is the effective channel gain, defined as  $|h_i|^2 \triangleq \chi_i \eta_i (r_i/r_0)^{-\alpha}$ ,  $r_i$  is the distance from  $U_i$  to the BS,  $\chi_i$  is a log-normal random variable modeling shadowing gain with standard deviation  $\sigma$  dB,  $\eta_i$  is small scale fading modeled by exponential random variable with parameter 1, and  $\alpha$  is the propagation loss factor. We further assume that the devices perform uplink power control such that the average received SNR from all devices at the BS are the same [13] and equals to  $\gamma_0$ . Such an assumption has been considered for code division multiple access (CDMA) in [14].

Time is divided into slots of duration  $\tau_s$  secs and total available bandwidth is  $W$  Hz. Each time slot is referred to

TABLE I  
NOTATION SUMMARY

Notation	Description
$r_0$	Cell radius
$W$	Total bandwidth in Hz
$\tau_s$	The duration of a resource block (RB)
$\tau$	The ratio of the duration of RA process in an RB
$h_i$	The effective channel gain of device $U_i$
$\gamma$	The reference SNR
$\gamma_0$	Received SNR at the BS after power control
$\lambda$	Average number of active devices in an RB
$\lambda_i$	Packet arrival rate at the $i^{th}$ device
$N$	Total number of devices
$N_s$	Total Number of RA preambles
$k$	Payload size of each MTC device
$\mathbf{b}^{(j)}$	Vector of information symbols of $U_j$
$\mathbf{G}^{(j)}$	Generator matrix of the AFC code used in $U_j$
$d_c$	AFC code degree
$\mathcal{W}_s$	Weight set of AFC
$D$	AFC weight set size (AFC maximum degree)
$\sigma_w^2$	Average weight power of an AFC
$N_d$	Number of delay groups
$t_i$	The delay constraint of the $i^{th}$ delay group
$p_i$	Access probability of the $i^{th}$ device
$L(\mathbf{s})$	Length of the random seed for devices with RA preamble $\mathbf{s}$
$\epsilon$	The probability that BS fails to detect random seeds
$\mathbb{Z}_{k_1}^{k_2}$	$\{k_1, k_1 + 1, \dots, k_2\} \subseteq \mathbb{Z}$ for $k_1 \leq k_2$

as a resource block (RB), where a fraction  $\tau$ ,  $0 < \tau < 1$ , of the resource block is reserved for the random access procedure as shown in Fig. 1. Similar to [13], we assume that the number of active devices in each RB of duration  $\tau_s$  is random and follows a Poisson process with rate  $\lambda$ . We also consider a contention-based random access strategy in M2M communications, where  $N_s$  different random access preambles, which are orthogonal frequency division multiplexing (OFDM) symbols in the LTE-Advanced standard [7], are selected in a random access attempt. The BS also assigns an access probability to each device according to its delay requirement. The access probabilities will be later optimized in this paper to satisfy the delay requirement of all devices. The notation used in this paper is summarized in Table I for quick reference.

It is important to note that in many M2M applications, M2M and human-to-human (H2H) communication coexists. H2H users have high priority to obtain a connection to transmit their data to the BS. Most of research work has considered that the contention-free random access is used for H2H users, that is the BS assigns a preamble and an RB to the H2H user in a timely manner [25]. In this work we only focus on M2M devices, where opportunistically contend for RBs through a contention-based random access, and H2H users are assumed to have access to the BS through the allocated RBs.

### B. Delay Model for MTC devices

In M2M communications, timing constraints for MTC devices typically range from 10 ms to several minutes due to the infrequent transmission features [1]. For instance, some M2M multimedia applications such as video surveillance systems with strict timing constraints have the delay requirement ranging from 10 ms to 40 ms. Such a diverse QoS requirements in M2M communications intensify the need for a practical channel assignment scheme to satisfy these requirements. Moreover, different MTC devices have diverse traffic patterns; i.e., some of them are regular and some have completely

random traffics. In Table II, we divide different M2M applications based on their delay requirements and traffic patterns, according to the 3rd Generation Partnership Project (3GPP) M2M standard in [1]. More specifically, we divide the MTC devices into  $N_d = 10$  groups based on their applications, where the maximum affordable delay in group  $i$  is denoted by  $t_i$ .

It is important to note that MTC devices with regular traffic have predictable activities and can be completely coordinated by the BS. This means that the BS will regularly assign an RB to the devices with regular activities and they can transmit their packets through the assigned RB. This can be easily obtained through several approaches in current LTE and LTE-A standards [22]. However, the devices with completely random activities should contend for RBs as allocating predetermined RBs to them is not efficient [1]. Therefore in this paper, we only consider MTC devices with random activities with irregular traffic patterns. This is the most challenging issue in M2M communications due to the ever increasing number of such devices.

Each active MTC device  $U_i$ ,  $i \in \mathbb{Z}_1^N$ , is assumed to have  $k$  information symbols to transmit within a maximum delay of  $t_j$ , where  $1 \leq j \leq N_d$ . That is, these  $k$  information symbols must be delivered to the BS within  $t_j$  seconds after  $U_i$  starts transmitting this message, i.e.,  $U_i$  belongs to the  $j^{th}$  delay group. We also assume that the  $i^{th}$  device has the packet arrival rate  $\lambda_i$ , that is in each RB,  $\lambda_i$  packets are generated at the  $i^{th}$  device, where each packet has to be delivered to the BS no later than  $t_j$  seconds after its generation. Let us assume that  $1/\lambda_i > t_j$ . Therefore, if a previous packet has been delivered at the BS no later than  $t_j$  seconds after its generation at the  $i^{th}$  device, the new packet will be immediately sent by the device and there is no need to buffer it. On the other hand, if  $1/\lambda_i < t_j$ , we set the delay requirement of the device to  $1/\lambda_i$  to avoid any further delay due to queuing. In other words, if the device has no memory to buffer the packets (i.e. each packet has to be successfully delivered at the BS before the generation of the next packet), the appropriate delay requirement for the device is  $\min\{1/\lambda_i, t_j\}$ . It is also important to note that for devices with limited buffer size, the device can change the access probability according to the number of packets in its buffer. Moreover, a packet in M2M communications has a small size (few bytes); therefore, we can assume that all the packets in the buffer of a MTC device can be transmitted in one RB [25]. Therefore In the rest of the paper, we only use the term delay requirement for each MTC device, and assume that the queuing delay has been taken into account when calculating the appropriate delay requirement of each device.

### III. THE PROPOSED RATELESS PROBABILISTIC MULTIPLE ACCESS FOR M2M COMMUNICATIONS

In this section, we develop an efficient multiple access scheme for M2M communications based on analog fountain codes. For this aim, we first briefly introduce AFC codes. Then, we present the rateless probabilistic multiple access for M2M communications, referred to as multiple access analog fountain coding (MA-AFC).

TABLE II  
SEVERAL M2M APPLICATIONS WITH THEIR DELAY CONSTRAINTS.

Group	Delay Constraint	Example of application	Traffic Type	Message Size
1	10 ms	Emergency Alarm [26]	Very Unlikely	Very small
2	20 ms	Intelligent transport system (ITS)	Regular/Irregular	Medium
3	40 ms	Video streaming	Regular (Streaming)	Large
4	100 ms	Control messages for devices [26]	Regular	Very small
5	500 ms	eHealth	Regular	Medium
6	1 s	Monitoring the Devices [26]	Regular	Small
7	10 s	eHealth	Regular/Irregular	Medium
8	100 s	eHealth	Regular/Irregular	Medium
9	500 s	Smart meters	Regular	Medium
10	>1000 s	Smart meters	Regular	Medium

#### A. Analog Fountain Codes

Analog fountain codes (AFC) were originally proposed in [15, 27] as an effective adaptive transmission scheme to approach the capacity of wireless channels. In AFC, the entire message of length  $k$  binary symbols is first modulated using binary phase shift keying (BPSK) modulation to obtain  $k$  modulated information symbols,  $b_i \in \{-1, 1\}$ , where  $i \in \mathbb{Z}_1^k$ . Then according to the predefined degree distribution function,  $d_c$  randomly selected modulated information symbols are linearly combined with real weight coefficients to generate one coded symbol. For simplicity, we assume that the code degree,  $d_c$  is fixed and weight coefficients are chosen from a finite weight set,  $\mathcal{W}_s$  with  $D$  positive real members, as follows:

$$\mathcal{W}_s = \{w_i \in \mathbb{R}^+ | i \in \mathbb{Z}_1^D\}, \quad (2)$$

where  $\mathbb{R}^+$  is the set of positive real numbers. Let us assume that  $\mathbf{b}$  is a BPSK modulated vector of dimension  $k \times 1$ , and  $\mathbf{G}$  is the generator matrix of dimension  $m \times k$ , then AFC coded symbols,  $\mathbf{c}$ , are generated as follows:

$$\mathbf{c} = \mathbf{G}\mathbf{b}, \quad (3)$$

where  $m$  is the number of coded symbols, only  $d_c$  elements of each row of matrix  $\mathbf{G}$  are nonzero, and each nonzero element of  $\mathbf{G}$  is randomly chosen from the weight set  $\mathcal{W}_s$ . To further enhance the performance, in [15] we proposed to use a high-rate precoder to encode the original data before applying the AFC code. To fully utilize the constellation plane, i.e., both in phase and quadrature phases, two consecutive coded symbols compose one signal as  $c_i + \sqrt{-1}c_{i+1}$ . We also use a standard belief propagation (BP) decoder to decode AFC codes. Further details of AFC encoding and decoding can be found in [15, 21].

#### B. Multiple Access AFC for M2M communications

The proposed probabilistic multiple access AFC (MA-AFC) scheme contains two steps. In the first step, namely *contention phase*, the random access requests are sent by the devices to the BS and a connection is established between the devices and the BS. In the second phase, namely *data transmission phase*, the actual payload data along with the device identity (ID) is transmitted by the devices to the BS.

1) *The contention phase (CP)*: To apply AFC codes to M2M communication systems and satisfy delay requirements of all devices, we need to slightly modify the random access procedure in LTE-A [7]. The BS first partitions  $N_s$  RA preambles into  $N_d \leq N_s$  subsets, according to delay constraints of MTC devices, and each group of RA preambles are assigned for one delay group. The BS then broadcasts this information to all devices. Let  $S_i$  denote the set of RA preambles allocated to the  $i^{th}$  delay group, then an MTC device with delay constraint  $t_i$  will randomly select an RA from  $S_i$  in the contention period. By detecting the RA preamble of the device, the BS will know which subset this RA preamble belongs to and thus know its delay requirement.

In the first step of CP, an MTC device with delay constraint  $t_i$  randomly selects an RA preamble from set  $S_i$  and transmits it to the BS. Let  $\mathbf{s} \in S_i$  denote the selected RA preamble, then, it is possible that more than one devices with the same delay constraint  $t_i$  have chosen  $\mathbf{s}$ . As we assume that all MTC devices perform power control, such that the received signals from different devices will have the same received power,  $\gamma_0$ , the received power at the BS for the RA preamble  $\mathbf{s}$  will be proportional to the number of devices which have selected RA preamble  $\mathbf{s}$ . Let  $\mathcal{U}(\mathbf{s})$  denote the set of MTC devices, which have selected the same RA preamble  $\mathbf{s}$ . Since we assume that the RA preambles are orthogonal, the received signal corresponding to RA preamble  $\mathbf{s}$  is given by:

$$\mathbf{y} = |\mathcal{U}(\mathbf{s})| \sqrt{\gamma_0} \mathbf{s} + \mathbf{z}, \quad (4)$$

where  $\mathbf{z}$  is the additive white Gaussian noise (AWGN) vector. The average received signal power at the BS can then be calculated as  $|\mathcal{U}(\mathbf{s})|^2 \gamma_0$ . Since the BS knows  $\gamma_0$ , it can easily estimate  $|\mathcal{U}(\mathbf{s})|$ .

In the second step of CP, the BS transmits a random access response (RAR) message including the number of devices which have selected each RA preamble, the allocated RB for each detected RA preamble, and timing information and optimized access probabilities. In the third step of CP, each device first calculates the length of a random seed based on the received information from the BS, and then generates a random seed with the determined length and sends it to the BS. We assume that the random seeds are orthogonal sequences. Once the BS and each device have shared the same seed for the random encoding structure of AFC, they can construct the same bipartite graph to perform the AFC encoding and decoding. Fig. 2 shows the proposed random access procedure for M2M communications.

Let  $L(\mathbf{s})$  denote the length of the random seed selected by devices which have selected the RA preamble  $\mathbf{s}$  in the first step of CP. The BS fails to uniquely detect these random seeds with the probability at most  $\epsilon$ , which is given by [13]:

$$\epsilon = \left(1 - \frac{1}{2^{L(\mathbf{s})} - 1}\right)^{|\mathcal{U}(\mathbf{s})| - 1}. \quad (5)$$

This is followed from the fact that each random seed can be detected at the BS if it has been generated by only one device. Thus, the minimum length of the random seed  $L(\mathbf{s})$  required to have a non-detectable probability lower than a predefined

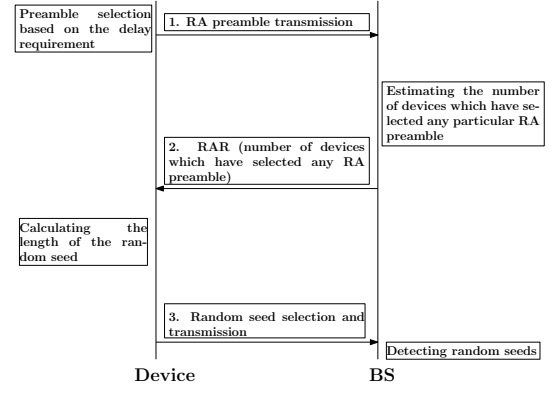


Fig. 2. The proposed random access procedure for M2M communications.

value  $\epsilon > 0$  is as follows:

$$L(\mathbf{s}) = \left\lceil \log_2 \left( 1 + \frac{1}{1 - (1 - \epsilon)^{\frac{1}{|\mathcal{U}(\mathbf{s})| - 1}}} \right) \right\rceil, \quad (6)$$

where  $\lceil \cdot \rceil$  is the ceil operator.

In the proposed RA scheme, even when two or more devices select the same preamble, they will be allocated with the same RB, thus they are transmitting in the same RB. This means that the proposed approach can support many users with a negligible collision probability, which can be controlled by the length of the random seed. Unlike the conventional RA schemes, the devices in our approach can send their random access requests whenever they have data to transmit and do not need to wait for several RA attempt to get access to the network. In fact, once the device sends a random access preamble to the BS, it will be allocated with an RB, no matter whether that preamble has been selected by other devices or not. This will significantly decrease the access delay.

2) *The data transmission phase*: After the contention phase, each active device and the BS have shared the same random seed. The devices which have selected the same RA preamble, are transmitting their AFC coded symbols in the same RB, in a probabilistic manner with the same access probability, but with different random seeds. Let  $p_j$  denote the access probability assigned to  $U_j$ . It is important to note that devices which have selected the same RA preamble will be assigned with the same access probability. In the data transmission phase, in each time instant  $\ell$ , device  $U_j$  generates a binary random number,  $I_{j,\ell}$ , which is one with probability  $p_j$ , i.e.,  $p(I_{j,\ell} = 1) = 1 - p(I_{j,\ell} = 0) = p_j$ . If this random number is one, then  $U_j$  generates an AFC coded symbol,  $u_\ell^{(j)}$  and sends it to the BS. The BS also generates the same binary random number as the one in  $U_j$  as they have shared the same random seed, thus it already knows in which time instants device  $U_j$  is sending a coded symbol. Using this random seed, both  $U_j$  and the BS can construct the same AFC code structure, so the BS is able to perform the decoding on the received coded symbols.

For simplicity, we assume that devices use the same code degree  $d_c$  and the same weight set  $\mathcal{W}_s$  to generate AFC coded symbols. Let  $\mathcal{U}$  denote the set of devices which are currently active and transmitting AFC coded symbols to the BS based on their access probabilities in the same RB. Then, the received

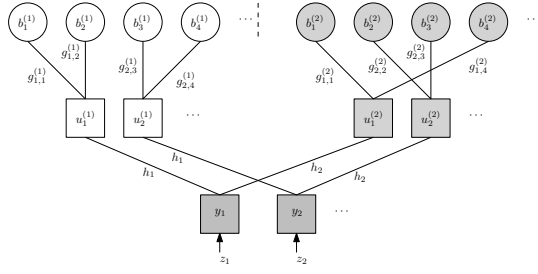


Fig. 3. Original AFC code graph at the BS.

signal at the BS in time instant  $\ell$  is given by

$$y_\ell = \sum_{j \in \mathcal{U}} I_{j,\ell} h_j u_\ell^{(j)} + z_\ell, \quad (7)$$

where  $z_\ell$  is the AWGN with variance  $\sigma_z^2$ . As can be seen in (7), the received signal at the BS is the noisy version of the sum of coded symbols of devices in  $\mathcal{U}$ . Since coded symbols of each device have been generated by a linear combination of modulated information symbols, the received signal at the BS can also be seen as a coded symbol of an analog fountain code, where modulated information symbols are from devices in  $\mathcal{U}$ . By using (3), (7) can be rewritten as follows:

$$y_\ell = \sum_{j \in \mathcal{U}} h_j I_{j,\ell} \sum_{r=1}^k g_{\ell,r}^{(j)} b_r^{(j)} + z_\ell, \quad (8)$$

where  $b_r^{(j)}$  is the  $r^{\text{th}}$  modulated information symbol of  $U_j$  and  $g_{\ell,r}^{(j)}$  is the respective weight coefficient. Fig. 3 and Fig. 4 show the original and equivalent bipartite graphs of the AFC codes at the BS, respectively, when there are only two devices in  $\mathcal{S}$  and  $d_c = 2$ . As shown in these figures, the equivalent bipartite graph can be considered as a bipartite graph of an equivalent AFC code where the information symbols are from both users. More specifically in the first time instant where  $I_{1,1} = I_{2,1} = 1$ , both devices transmit coded symbols to the BS. But, in the second and third time instant, only one of the devices transmits a coded symbols to the BS, i.e.,  $I_{1,2} = I_{2,3} = 1$  and  $I_{1,3} = I_{2,2} = 0$ . This is because of the fact that the devices transmit coded symbols based on their access probabilities. In fact, a device  $U_j$  with access probability  $p_j$  transmit a coded symbols with probability  $p_j$ ; otherwise it remains silent.

Let us define  $X_{j,r}^{(\ell)} \triangleq I_{j,\ell} g_{\ell,r}^{(j)} b_r^{(j)}$ . Then  $X_{j,r}^{(\ell)}$  is a random variable with the following probability distribution:

$$p(X_{j,r}^{(\ell)} = s) = \begin{cases} 1 - \frac{p_j d_c}{k} & \text{if } s = 0, \\ \frac{p_j d_c}{2kD} & \text{if } s \in \mathcal{W}_s \text{ or } -s \in \mathcal{W}_s. \end{cases}$$

Thus, the mean and variance of  $X_{j,r}^{(\ell)}$  can be calculated as follows:

$$m_j = \mathbb{E}[X_{j,r}^{(\ell)}] = \sum_{s \in \mathcal{W}_s} \frac{p_j d_c}{2kD} (s - s) = 0, \quad (9)$$

$$\sigma_j^2 = \mathbb{E}[X_{j,r}^{(\ell)2}] = \sum_{s \in \mathcal{W}_s} \frac{p_j d_c}{kD} s^2 = \frac{1}{k} p_j d_c \sigma_w^2, \quad (10)$$

where  $\sigma_w^2 = \frac{1}{D} \sum_{i=1}^D w_i^2$ . Since in M2M communications the number of devices are very large and  $X_{j,l}$ 's are independent random variables, according to the central limit theorem  $\sum_{j \in \mathcal{U}} \sum_{l=1}^k X_{j,l}$  will approximately follow a Gaussian distri-

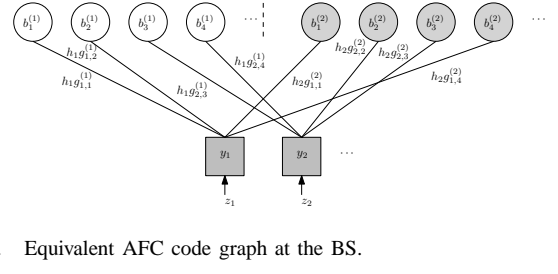


Fig. 4. Equivalent AFC code graph at the BS.

bution with a zero mean and the following variance:

$$\sigma_Y^2 = d_c \sigma_w^2 \sum_{j \in \mathcal{U}} p_j |h_j|^2. \quad (11)$$

In other words, the received signals at the BS are normally distributed around the origin with average power  $\sigma_Y^2$ . As the received coded symbols at the BS can be represented as coded symbols of an equivalent AFC code, the standard BP decoding algorithm can be applied to jointly decode all active devices at the BS.

#### IV. ANALYSIS OF THE RATELESS MULTIPLE ACCESS FOR M2M COMMUNICATIONS

In this section, we analyze the performance of the proposed MA-AFC scheme by using a density evolution based approach. These results will be used as a basis to optimize the proposed rateless multiple access for M2M communications.

As the BS determines the allocation of RBs to the devices, we assume a general case, where the number of active devices allocated to each RB is modeled by a Poisson random variable with parameter  $\lambda$ . We also assume that each active device belongs to a delay group with a uniform probability. This model enables us to have a fair comparison with the throughput limit obtained in [13]. Thus, in the rest of the paper, we only consider one RB and optimize the access probabilities for the proposed scheme to simultaneously satisfy delay requirements for all devices.

##### A. Optimum Coordinated Multiple Access

Before analyzing the proposed multiple access scheme, let us first investigate the optimum coordinated multiple access, which will be used as a baseline for the proposed probabilistic multiple access AFC scheme.

We assume that the base station has already identified the devices which are transmitting in the same RB. The BS also knows all the channel information and the devices are transmitting with the same power. As shown in [13], the maximum common rate that can be achieved by  $N$  devices transmitting in a typical RB with ordered effective channel gains  $\{h_i\}_1^N$  and reference SNR  $\gamma$  is given by:

$$\mathcal{R}_c = \min_{j \in \mathcal{Z}_1^N} \frac{1}{j} \log_2 \left( 1 + \gamma \sum_{\ell=1}^j |h_\ell|^2 \right) \text{ bps/Hz}, \quad (12)$$

which is simply obtained from the MAC capacity region when the transmission rate of all the devices are the same. This can be used as an upper bound on the average system common rate for the random access and later we will show how the proposed MA-AFC code can approach this limit. More specifically,

when the devices perform power control in a way to have the same received SNR  $\gamma_0$  at the BS, the common rate is given by:

$$\mathcal{R}_c = \frac{1}{N} \log_2(1 + N\gamma_0) \text{ bps/Hz.} \quad (13)$$

Therefore, the average system throughput increases with the number of devices with a logarithmic slope. The maximum payload size  $L_{opt}$  which can be successfully transmitted by each device in a resource block of duration  $\tau_s$  and bandwidth  $W$  is then given by:

$$L_{opt} = \frac{W\tau_s}{N} \log_2(1 + N\gamma_0) \text{ bits,} \quad (14)$$

which is obtained without considering the overhead for the contention phase. In fact, (14) provides an upper bound on the maximum payload size which can be delivered from each active device at the BS.

### B. Asymptotic Performance Analysis of Multiple Access AFC based on the BP Decoding

A commonly used analytical tool for analyzing a BP decoder is density evolution [28], which calculates the evolutions of the message passing in the iterative decoding process. In this paper, we focus on the density evolution analysis in the asymptotic case, when the number of variable and check nodes go to infinity.

Let us refer to information symbols of device  $U_i$  as Type- $X_i$  variable nodes in the bipartite graph of MA-AFC codes, for  $i \in \mathbb{Z}_1^N$ . Each received coded symbol at the BS may connect to modulated information symbols of various sets of devices. A simple way to represent coded symbols at the BS is to divide them into different types according to their connections to different sets of devices. For simplicity, we represent each non-empty set of devices by a vector  $\mathbf{v}$  of dimension  $N$ , where its  $i^{th}$  entry,  $v_i$ , is 1 if  $U_i$  belongs to the set; otherwise, it is zero. Here, we refer to coded symbols which are connected to the modulated information symbols of set  $\mathbf{v}$  as Type- $\mathbf{v}$  check nodes. It is easy to show that the probability that a coded symbol is of Type- $\mathbf{v}$ ,  $q_{\mathbf{v}}$ , is given by:

$$q_{\mathbf{v}} = \prod_{i=1}^N p_i^{v_i} (1 - p_i)^{1-v_i}, \quad (15)$$

and there are in total  $2^N - 1$  various types of check nodes due to the fact that the total number of nonempty subsets of devices is  $2^N - 1$ .

Since there are  $2^N - 1$  types of variable and check nodes, we need to analyze the message between each specific types of variable and check nodes. In [29, 30], an AND-OR tree analytical method was proposed to analyze the decoding error probability of LT codes [31] with different types of variable and check nodes for erasure channels. However, the AND-OR tree approach cannot be applied directly to wireless channels. Here, we extend the conventional density evolution analysis to the case that there are multiple types of variable and check nodes.

*Lemma 1:* Let us consider an M2M system consisting of  $N$  MTC devices wanting to simultaneously transmit their messages to a common BS. Each device  $U_i$  is assigned with

an access probability  $p_i$  and code degree  $d_i$ . Let  $L_{X_i \rightarrow \mathbf{v}}^{(t)}$  denote the message passed from a Type- $X_i$  variable node to a Type- $\mathbf{v}$  check node in the  $t^{th}$  iteration of the BP decoding algorithm at the BS. Then  $L_{X_i \rightarrow \mathbf{v}}^{(t)}$  can be approximated by a normal distribution with mean  $m_i^{(t)}$  and variance  $2m_i^{(t)}$ , where  $m_i^{(t)}$  can be calculated as follows:

$$m_i^{(t)} = \gamma |h_i|^2 \sigma_w^2 d_{v_i} \sum_{\mathbf{v}: v_i=1} \frac{2}{1 + \sigma_{\mathbf{v}}^2} q_{\mathbf{v}}, \quad (16)$$

where  $d_{v_i} = md_i/k$ ,  $\sigma_w^2 = \frac{1}{D} \sum_{i=1}^D w_i^2$ ,

$$\sigma_{\mathbf{v}}^2 = \sum_{i=1}^N \gamma |h_i|^2 d_i v_i \sigma_w^2 \Omega(m_i^{(t)}), \quad (17)$$

and  $\Omega(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} (1 - \tanh(x - y\sqrt{x})) e^{-\frac{y^2}{2}} dy$ .

The proof of this lemma is provided in Appendix A. The BER for the  $i^{th}$  device after  $t$  iterations of the BP decoder, denoted by  $P_{e,i}^{(t)}$ , can then be calculated as follows:

$$P_{e,i}^{(t)} = Q\left(\sqrt{m_i^{(t)}}\right), \quad (18)$$

where  $Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} e^{-z^2/2} dz$ .

### C. Discussions on the asymptotic MA-AFC decoding performance at high SNR

As stated before, in MA-AFC, the BS receives the sum of the coded symbols generated from active users. The BS thus tries to decode each device's data from the received signals by performing the BP decoding. Now let us consider the asymptotic performance of MA-AFC as SNR goes to infinity, thus ignoring the noise. The decoding problem in this case is equivalent to solving a system of binary linear equations, where the variables are the information symbols of all active devices. More specifically, the BS requires to solve the following system of binary linear equations to find  $b_l^{(j)}$  for  $j \in \mathbb{Z}_1^N$  and  $l \in \mathbb{Z}_1^k$ , given the value of  $y_i$  and matrices  $\mathbf{G}^{(j)}$ :

$$y_i = \sum_{j=1}^N \sum_{l=1}^k I_{jg_{il}^{(j)}} b_l^{(j)}. \quad (19)$$

Let us define matrix  $\mathbf{G} \triangleq [\mathbf{G}^{(1)} | \mathbf{G}^{(2)} | \dots | \mathbf{G}^{(N)}]$  and binary vector  $\mathbf{b} \triangleq [\mathbf{b}^{(1)} | \mathbf{b}^{(2)} | \dots | \mathbf{b}^{(N)}]$ . It is clear that  $\mathbf{G}$  is the generator matrix of the equivalent AFC code at the BS and we have  $\mathbf{Y} = \mathbf{G}\mathbf{b}'$ . Since we consider that different devices select the weight coefficients from the same set  $\mathcal{W}_s$ , it is possible that two different devices select the same weight coefficients for their transmitted symbols at the same time. More specifically, at time instant  $i$ , the probability that coded symbols of device  $U_\ell$  and  $U_m$  have at least one common weight coefficient, denoted by  $p_{\ell,m}$ , is given by

$$p_{\ell,m} = 1 - \binom{D - d_\ell}{d_m} / \binom{D}{d_m}, \quad (20)$$

where  $D$  is the weight set size and we assume that  $D \gg d_\ell$  and  $D \gg d_m$ . It is clear that when at least two columns of  $\mathbf{G}$  are exactly the same, then (19) does not have a unique solution. The following lemma gives the probability that at least two column of matrix  $\mathbf{G}$  are exactly the same.

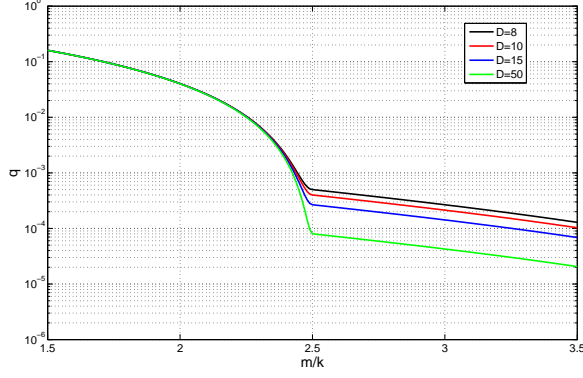


Fig. 5. The probability that at least two columns of the generator matrix at BS are the same versus the ratio of the number of coded symbols and that of information symbols for different weight set sizes. Each device has the access probability of 0.1 and code degree of 4.

**Lemma 2:** Let us consider that  $m$  coded symbols are received at the BS and each information symbol of each device is connected to at least one coded symbol at the destination, i.e., the set of non-zero elements in each column of  $\mathbf{G}$  is not empty. Moreover, we assume that different devices have the same access probability  $p$  and code degree  $d_c$ . Let  $q$  denote the probability that two columns of  $\mathbf{G}$  are exactly the same, then  $q$  can be calculated as follows:

$$q = \frac{\frac{(1+\beta-d_v)^2(m-d_v+1)}{Dd_v} + (d_v - \alpha)^2}{D^{d_v-1} \binom{m}{d_v-1}}, \quad (21)$$

where  $\beta = mp d_c / k$  and  $d_v = \lceil \beta \rceil$ .

The proof of this lemma is provided in Appendix B. From this lemma, as can be also seen in Fig. 5, even when the weight set size is small, we can always uniquely decode all devices' information symbols by transmitting more coded symbols. Moreover, when  $D$  is small, devices are required to transmit more coded symbols to guarantee that the BS can uniquely decode all devices information symbols, which clearly decrease the overall system throughput. To achieve higher throughput one may consider larger code degree  $d_c$ , which on the other side significantly increases the decoding complexity at the BS that exponentially increases with  $d_c$ .

#### D. How much overhead is required in the contention phase?

In the first phase of the contention phase, each active device sends a RA preamble to the BS. Let  $L_{cp,1}$  denote the length of the RA preamble. In the second phase, the BS sends the information about the number of devices selected a particular RA preamble. The average number of devices which have selected the same RA preamble is  $N/N_s$ , where  $N$  is the number of active devices and  $N_s$  is the number of RA preambles. These information can be broadcasted by the BS by using at least  $L_{cp,2} = N_s \lceil \log_2(N/N_s) \rceil$  information bits. In the third phase of the contention phase, the users will select a random seed of length  $L(s)$  which is calculated by (6), for a given value of  $\epsilon$  and a given RA preamble  $s$ . Thus, at most  $L_{cp,3} = N_s L(s)$  information bits have to be sent to the BS in the third phase of the contention period. Therefore, the total

overhead can be calculated as follows:

$$\begin{aligned} L_{ov} &= L_{cp,1} + L_{cp,2} + L_{cp,3} \\ &= L_{cp,1} + N_s \lceil \log_2(N/N_s) \rceil \\ &\quad + N_s \left\lceil \log_2 \left( 1 + \frac{1}{1 - (1-\epsilon)^{\frac{1}{(N/N_s)-1}}} \right) \right\rceil \end{aligned} \quad (22)$$

As the BS allocates RBs to devices and devices with the same RA preamble will transmit at the same RB, each device then knows at which RB it should transmit its message based on its delay requirement. In this case, the amount of overhead can be further reduced to

$$\begin{aligned} L_{ov} &= L_{cp,1} + \lceil \log_2(N/N_s) \rceil \\ &\quad + \left\lceil \log_2 \left( 1 + \frac{1}{1 - (1-\epsilon)^{\frac{1}{(N/N_s)-1}}} \right) \right\rceil \end{aligned} \quad (23)$$

Let  $k$  denote the message length of each device and  $\mathcal{R}_c$  denote the achievable common rate for a given arrival rate  $\lambda$  using the proposed probabilistic multiple access approach. Then, this message can be successfully transmitted to the BS within a resource block of duration  $\tau_s$  and bandwidth  $W$  if the following condition is satisfied:

$$k + L_{ov} \leq \tau_s W \mathcal{R}_c. \quad (24)$$

In other words, the maximum message size (in bits) which can be supported by the proposed approach in an RB can be approximated by:

$$k_{max} = \max\{\tau_s W \mathcal{R}_c - L_{ov}, 0\}. \quad (25)$$

#### V. OPTIMIZATION OF ACCESS PROBABILITIES WITH DELAY GUARANTEES

In this section, we formulate an optimization problem to simultaneously satisfy the delay requirements of all the devices. The access probability and code degree for each device will be optimized in a way that delay constraints of all devices are satisfied. Here we minimize the average degree of the received signal at the BS to further reduce the complexity of the decoder. Therefore, the BS should perform the following optimization process:

$$\min \sum_{i=1}^N p_i d_i$$

subject to (a).  $Q\left(\sqrt{m_i^{(\infty)}}\right) < \delta$ , (b).  $1 \leq d_i \leq D$ , and (c).  $0 < p_i \leq 1$ , for  $i \in \mathbb{Z}_i^N$ , where  $\delta > 0$  is the target bit error rate,  $m_i^{(\infty)} = \lim_{t \rightarrow \infty} m_i^{(t)}$ , and

$$m_i^{(t)} = \frac{\gamma |h_i|^2 \sigma_w^2 d_i t_i}{k} \sum_{v_i=1}^2 \frac{2}{1 + \sum_{j=1}^N \gamma |h_j|^2 d_j v_j \sigma_w^2 \Omega\left(m_j^{(t-1)}\right)^{q_v}}.$$

It is important to note that  $\Omega(x)$  is very close to zero for large values of  $x$ , so we can approximate (16) as follows:

$$m_i^{(t)} \approx \frac{2\gamma |h_i|^2 \sigma_w^2 d_i t_i p_i}{k}, \quad (26)$$

for large  $x$  values. This means that the device  $U_m$  which has the highest value of  $|h_m|^2 d_m p_m$ , will be decoded sooner than



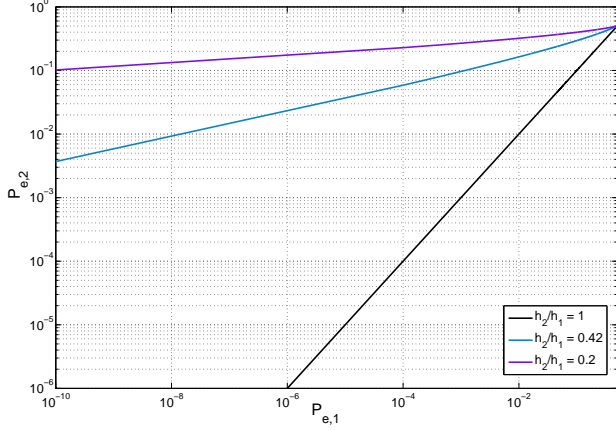


Fig. 6. BER of device  $U_2$  versus the BER of device  $U_1$  for the case that both the users have the complete access to the channel and  $d_1 = d_2$ .

other devices at the BS. Let  $P_{e,i}$  denote the bit error rate of device  $U_i$  at the BS. Thus, by using (26) we have:

$$\frac{m_i^{(t)}}{m_j^{(t)}} \approx \frac{|h_i|^2 d_i p_i}{|h_j|^2 d_j p_j}, \quad (27)$$

where we assume that  $\sigma_v^2$  is constant which is valid when the number of devices is very large. Therefore, according to (18) we have:

$$P_{e,i}^{(t)} \approx Q \left( \sqrt{[Q^{-1}(P_{e,j}^{(t)})]^2 \frac{|h_i|^2 d_i p_i}{|h_j|^2 d_j p_j}} \right). \quad (28)$$

More specifically, when the devices perform power control to have the same received SNR at the BS and use the same code degree  $d_c$ , then (28) can be further simplified as follows:

$$P_{e,i}^{(t)} \approx Q \left( \sqrt{[Q^{-1}(P_{e,j}^{(t)})]^2 \frac{p_i}{p_j}} \right), \quad (29)$$

which clearly shows that the device which has a higher access probability is decoded sooner than those with lower access probabilities. Fig. 6 shows the BER for device  $U_2$  versus the BER of device  $U_1$  for the case that both users have complete access to the channel, i.e.,  $p_1 = p_2 = 1$ . As can be seen in this figure, when the channel gains are the same, the BER performance is the same. Moreover, when the ratio of the channel gain of  $U_1$  and that of  $U_2$  decreases, the BER of  $U_2$  also decreases.

By using (29), we can simplify the optimization problem as follows, assuming that all the devices use the same code degree  $d_c$  and perform the power control to have the same received SNR  $\gamma_0$  at the BS.

$$\min \sum_{i=1}^N p_i$$

subject to (a).  $p_i \geq \frac{k[Q^{-1}(\delta)]^2}{2\gamma_0 d_c t_i}$ , and (b).  $0 < p_i \leq 1$ , for  $i \in \mathbb{Z}_i^N$ . Therefore, the optimized access probabilities can be found as follows:

$$p_{i,opt} = \min \left\{ \frac{k[Q^{-1}(\delta)]^2}{2d_c \gamma_0 t_i}, 1 \right\}. \quad (30)$$

According to (30) we have  $p_{i,opt} \propto 1/t_i$ . Thus, if device  $U_i$  has the access probability of  $p_i$ , then the access probability

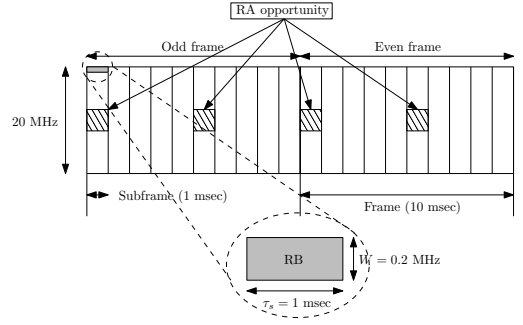


Fig. 7. Physical layer configuration in LTE-A [22].

for  $U_j$  will be simply  $p_j = t_i/t_j$ , which further simplifies the optimization process at the BS. Furthermore, as the devices which have selected the same RA preamble will transmit in the same RB with the same access probability, the BS does not need to calculate the optimal access probability for every device. Instead, it calculated the access probabilities for each group of devices which have selected the same RB, and allocate the same access probability to devices within the same delay group.

## VI. SIMULATION RESULTS

In this section, we investigate the performance of the proposed MA-AFC scheme in M2M communication systems. We first compare our proposed MA-AFC approach with the conventional massive RA technique in the LTE-A standard [32]. Like [22], as shown in Fig. 7 we assume that time is divided into time frames of length 10 msec and the total system bandwidth is 20 MHz. Each frame is further divided into 10 subframes of length 1 msec. A subframe consists of 100 RBs, each with bandwidth  $W = 0.2$  MHz. We also assume a total number of  $N_s = 60$  preambles and for simplicity we assume several configurations of RAs, which includes 1 RA attempt per two time frames, 1 RA access attempt per time frame, and 2 RA attempts per time frame. Each successful device will then be allocated with 16 RBs for the data transmission in next subframes. We assume that the packet length of devices is small (few bytes), thus a device can successfully transmit its packet within one time frame. The conventional random access approach, which is currently used in the LTE-A standard uses backoff approach [7]. In this approach, when a random access attempt is unsuccessful, the device will wait for a random time before transmitting the new access request. More specifically, before the  $i^{th}$  random access request, the device will wait for time  $t_{ra}^{(i)}$ , where  $t_{ra}^{(i)}$  is randomly drawn from range  $[0, X_i - 1]$ . The device can send a maximum number of  $N_{ra}$  RA request using this backoff approach. More specifically, we assume maximum number of RA attempts  $N_{ra} = 10$  and backoff indexes  $X_i \in \{10, 20, 30, 40, 60, 80, 120, 160, 240, 320, 480, 960\}$  [22].

For the comparison, we use two metrics, blocking probability and average access delay. The blocking probability is defined as the ratio of the MTC devices failing to access the BS due to exceeding their maximum allowed random access attempts to the total number of MTC devices. The average access delay is defined as the average time elapsed from the time instant when a device sends the first access request until

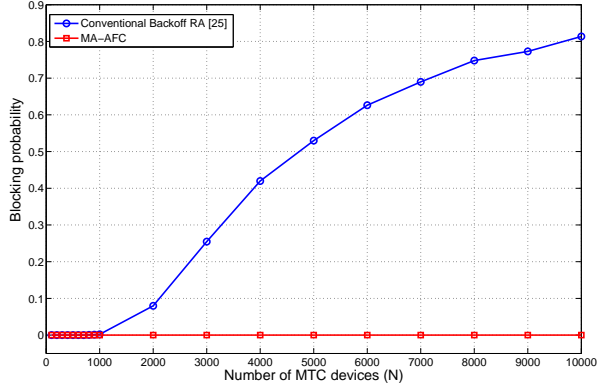


Fig. 8. The blocking probability for MTC devices versus the number of devices. The total number of preambles is  $N_s = 60$ . All device in the MA-AFC case use the same code degree 8.

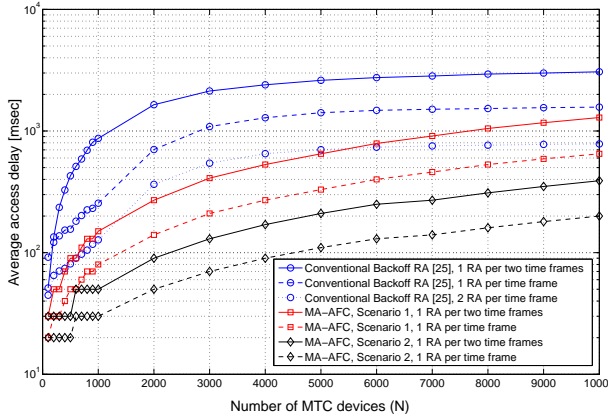


Fig. 9. The average access delay for MTC devices versus the number of devices. The total number of preambles is  $N_s = 60$ . The access probability for MTC devices in Scenario 1 is  $N_s/N$  and in Scenario 2 is  $10N_s/N$ . All device in the MA-AFC case use the same code degree 8.

that it succeeds. Fig. 8 shows the blocking probability for both the proposed and conventional approaches. As can be seen in this figure, the blocking probability for the conventional approach is very high especially when the number of devices is large. However, devices in the proposed scheme can access to the BS with a very small blocking probability, due to the fact that the devices can transmit simultaneously and they do not need to wait until they select separate preambles.

Fig. 9 shows the average access delay versus the number of devices. As can be seen in this figure, the proposed approach outperforms the conventional approach and have roughly 70% less access delay. In Fig. 9, we considered two scenarios for MA-AFC. The access probability for all devices in Scenario 1 and Scenario 2 is set to  $N_s/N$  and  $10N_s/N$ , respectively. It is important to note that the proposed approach in Scenario 2 has a lower access delay due to the fact that more devices can simultaneously transmit but it has higher complexity at the BS due to larger degrees of equivalent AFC code at the BS.

We then simulate our proposed approach within one RB, and compare the results with the fundamental throughput limits in [13]. For this aim, we assume that the reference SNR is  $\gamma = 0$  dB. This is equivalent to the case that a device transmits with 10 dBm (10mW) power over 1MHz bandwidth, a noise power

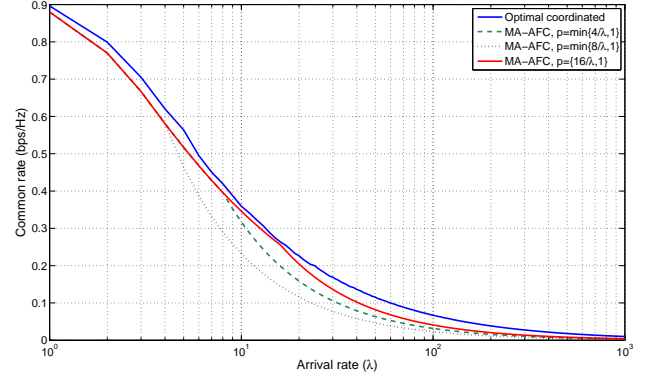


Fig. 10. Maximum achievable common rate versus the arrival rate for the case that the devices perform power control to have the same received SNR of 0 dB at the BS.

spectral density of -174dBm/Hz, a receiver noise figure of 5dB, a receiver antenna gain of 14dB, a 3.76 pathloss exponent, a 128dB pathloss intercept at 1000m, and a cell radius of 1360m [13]. We also assume that  $\tau_s = 1$  sec.

Let us first investigate the maximum common rate which can be achieved by the proposed probabilistic MA-AFC approach in the contention-free case. Fig. 10 shows the maximum common rate versus the arrival rate for the case that the devices perform the power control to have the same received SNR of 0 dB at the BS. We consider three cases for the MA-AFC approach with different access probabilities. As can be seen in this figure, the achievable rate of the proposed approach is very close to the optimal coordinated multiple access [13]. It is important to note that in the proposed approach we design the access probabilities in a way that the number of devices which simultaneously transmits is 4, 6, and 8, respectively, in order to minimize the decoding complexity at the BS. Fig. 10 shows that even with a small access probability, the achievable common rate of the proposed approach is very close to the optimal coordinated approach, especially when the arrival rate is very large.

In Fig. 11, we show the maximum payload size which can be supported by the proposed approach in a resource block of duration  $\tau_s = 1$  sec and bandwidth  $W = 10$  kHz for different arrival rates. An upper bound on the maximum payload size which has been obtained from (14) is also shown in this figure. As can be seen in this figure, as the arrival rate increases, the payload size which can be supported decreases. This is because that the common rate is a decreasing function of the arrival rate (Fig. 10). Moreover, as the arrival rate increases, the overhead calculated in (23) increases, which reduces the supported payload size. Also, by increasing the access probability, a higher common rate can be achieved, which leads to higher supported payload sizes.

Fig. 12 shows the average BER versus the number of received coded symbols at the BS when the code degree for all devices is  $d_c = 8$  and the reference SNR is  $\gamma = 30$  dB. As can be seen in this figure, by increasing the number of devices, the BS requires more coded symbols, in order to decode all devices' information symbols. Moreover, by increasing the number of devices, the average number of required coded symbols per device remains relatively constant to achieve a certain

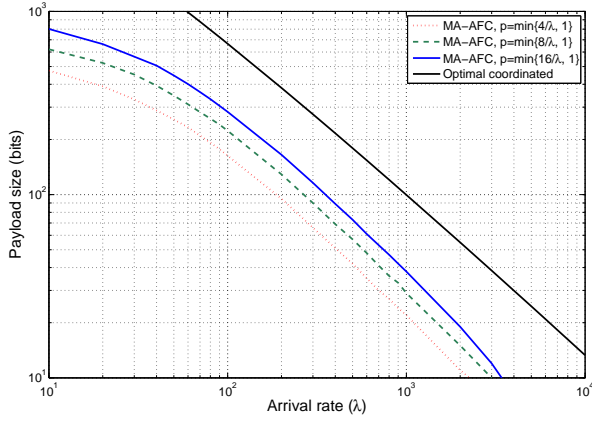


Fig. 11. Maximum supported payload size versus the arrival rate. Devices perform power control to have the same received SNR of 0 dB at the BS. The devices use the same code degree of 8 and the same weight set  $\mathcal{W}_s$ .

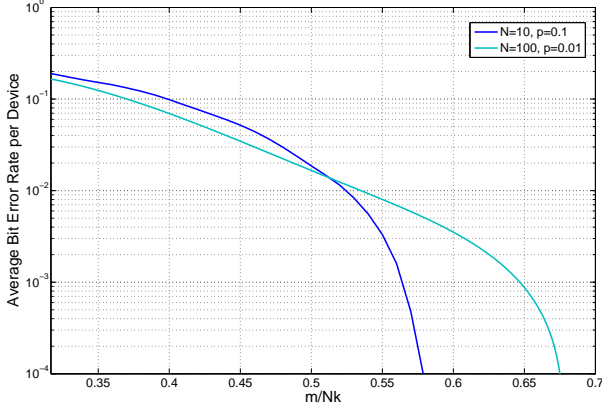


Fig. 12. BER versus the ratio of the number of coded symbols and the total number of information symbols.  $m$  is the number of coded symbols,  $N$  is the number of devices, and  $k$  is the length of the message for each device.

BER. This indicates that the proposed scheme can provide a certain level of BER at the BS for each device even with a large number of devices and the decoding performance is not significantly degraded by increasing the number of devices.

We now consider the case that devices have different delay requirements. We assume that the payload size of each device is 1 kb and 100 MTC devices belongs to each specific delay group, as shown in Table II. By using the optimization approach in Section IV, we can derive the optimized access probability for each group. Fig. 13 shows the average BER versus the delay for devices in each group of MTC devices specified in Table II, when all the devices has the same code degree  $d_c = 8$ . As can be seen in this figure, the delay requirement for devices in various groups have been simultaneously satisfied.

## VII. CONCLUSIONS

In this paper, we proposed a novel rateless multiple access scheme to support a large number of devices in M2M communications. This is achieved by representing the multiple access process by a capacity approaching analog fountain code to minimize the access delay while achieving a higher throughput. In the proposed scheme, devices are allowed

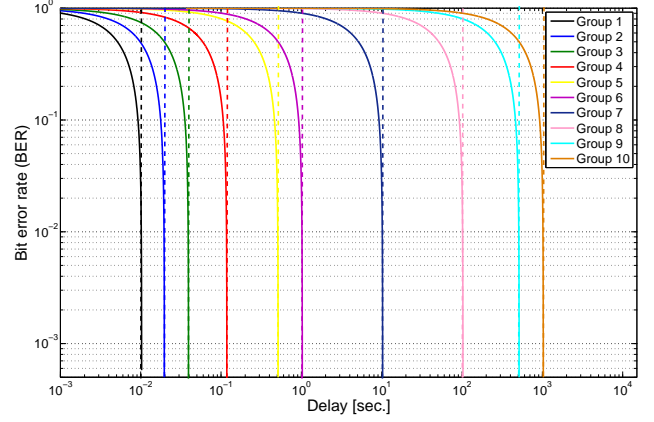


Fig. 13. The average bit error rate for different groups of devices classified by delay requirements. The total number of devices is  $N = 1000$ , each group has 100 devices; the devices use the same code degree  $d_c = 8$ , access probabilities are calculated from (30) and  $\gamma = 0$  dB. Dashed lines show the delay requirement for different groups of devices.

to transmit at the same time and the same channel, thus forming an equivalent analog fountain code at the base station (BS). The BS then performs joint decoding to recover all devices' messages. We further analyzed the proposed multiple access analog fountain code by using the density evolution technique and formulated an optimization problem to find the optimum access probability for each device to satisfy the delay requirements of MTC devices. Simulation results showed that the proposed scheme closely approaches the optimal common-rate performance of the multiple access channel, even for a very large number of devices. Moreover, the proposed scheme can satisfy the delay requirements for all the devices with a relatively low decoding complexity, which increases linearly with the number of devices. We also proposed a simple contention-based strategy and formulated the average overhead of the proposed scheme. Our novel random access and rateless multiple access schemes also achieved a significantly lower access delay and blocking probability compared to existing massive random access schemes for cellular-based M2M communications.

## APPENDIX A PROOF OF LEMMA 1

Let us consider the equivalent AFC code at the BS, where the equivalent generator matrix and the message vector are denoted by  $\mathbf{G}$  and  $\mathbf{b}$ , respectively. We first define  $Y_j \triangleq \sum_{r' \in \mathcal{M}(j) \setminus r} g_{j,r'} b_{r'}$ , where  $\mathcal{M}(j) \setminus r$  is the set of all variable nodes connected to check node  $y_j$  except variable node  $b_r$ . In [33], it has been shown that the average log likelihood ratio (LLR) passed from check node  $y_j$  to variable node  $b_r$  in the  $t^{\text{th}}$  iteration of the BP algorithm,  $L_{j \rightarrow r}^{(t)}$ , can be approximated as follows:

$$\mathbb{E} \left\{ L_{j \rightarrow r}^{(t)} \right\} = 2g_{j,r} \int_{-\infty}^{+\infty} \frac{f_1(y)}{f_0(y)} dy,$$

where  $f_m(y)$  is defined as follows for  $m \in \{0, 1\}$ :

$$f_m(y) = \sum_{r' \in \mathcal{M}(j) \setminus r} \frac{e^{-\frac{1}{2}(y-Y_j)^2}}{\sqrt{2\pi}} (y-Y_j)^m \prod_{r' \in \mathcal{M}(j) \setminus r} p_X(b_{r'}).$$

As shown in [33], when the number of variable and check nodes go to infinity,  $\int_{-\infty}^{+\infty} \frac{f_1^2(y)}{f_0(y)} dy$  tends to  $\frac{1}{1+\sigma_Y^2}$ , where  $\sigma_Y^2$  is the variance of  $Y_j$ . As shown in [33], in the asymptotic case the message passed from a variable to a check node is normally distributed and we can find the mean and variance of the message passed from a Type- $X_i$  variable node to a Type- $v$  check node, denoted by  $L_{X_i \rightarrow v}^{(t)}$ , as follows:

$$\mathbb{E} \left\{ L_{X_i \rightarrow v}^{(t)} \right\} = \frac{2}{1 + \sigma_v^2} \sum_{\ell' \in \mathcal{N}(r) \setminus \ell} g_{\ell', r}^2,$$

$$\text{var} \left\{ L_{X_i \rightarrow v}^{(t)} \right\} = 2 \mathbb{E} \left\{ L_{r \rightarrow \ell}^{(t)} \right\},$$

where  $\mathcal{N}(X_i) \setminus \ell$  is the set of all check nodes connected to a Type- $X_i$  variable node except check node  $y_\ell$ . As we assume that the number of variable and check nodes go to infinity, according to the law of large number, for large code degree  $d_i$ ,  $\sum_{\ell' \in \mathcal{N}(X_i) \setminus \ell} g_{\ell', X_i}^2$  can be approximated by  $\sigma_w^2 m d_i / k$ , where  $m$  is the number of coded symbols. As a Type- $X_i$  variable node is connected to a Type- $v$  check node with probability  $q_v$ , then the average total LLR value for a Type- $X_i$  can be found as follow:

$$m_i^{(t)} = \gamma |h_i|^2 \sigma_w^2 d v_i \sum_{v=1}^V \frac{2}{1 + \sigma_v^2} q_v,$$

where  $\sigma_v^2 = \sum_{i=1}^N d_i \gamma |h_i|^2 \sigma_w^2 v_i \Omega(m_i^{(t)})$  and  $\Omega(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} (1 - \tanh(x - y\sqrt{x})) e^{-\frac{y^2}{2}} dy$  [33]. This completes the proof.

## APPENDIX B PROOF OF LEMMA 2

To generate each AFC coded symbol,  $d_c$  information symbols are selected from those with minimum degrees, so after receiving  $m$  coded symbols at the BS, the degree of each information symbol will be either  $d_v$  and  $d_v - 1$ , with probabilities  $1 + \beta - d_v$  and  $d_v - \beta$ , respectively. Moreover, two columns of  $\mathbf{G}$  are the same if the number of their nonzero elements are the same. The probabilities that two columns of  $\mathbf{G}$  have the same degree  $d_v$  and  $d_v - 1$ , are  $(1 + \beta - d_v)^2$  and  $(d_v - \beta)^2$ , respectively. Furthermore, the probability that the nonzero elements of the two columns of the same degree are the same is  $1/\binom{m}{a} D^a$ , where  $a$  is either  $d_v$  or  $d_v - 1$ . This arises from the fact that each nonzero element of each column is randomly chosen from  $m$  possible positions, and the value of each nonzero element is randomly selected from  $D$  possible weights. This completes the proof.

## REFERENCES

- [1] S.-Y. Lien, K.-C. Chen, and Y. Lin, "Toward ubiquitous massive accesses in 3GPP machine-to-machine communications," *IEEE Commun. Mag.*, vol. 49, no. 4, pp. 66–74, 2011.
- [2] G. Wu, S. Talwar, K. Johansson, N. Himayat, and K. Johnson, "M2M: From mobile to embedded internet," *IEEE Commun. Mag.*, vol. 49, no. 4, pp. 36–43, 2011.
- [3] H. Dhillon, H. Huang, H. Viswanathan, and R. Valenzuela, "On resource allocation for machine-to-machine (M2M) communications in cellular networks," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, 2012, pp. 1638–1643.
- [4] K. S. Ko, M. J. Kim, K. Y. Bae, D. K. Sung, J. H. Kim, and J. Y. Ahn, "A novel random access for fixed-location machine-to-machine communications in OFDMA based systems," *IEEE Commun. Lett.*, vol. 16, no. 9, pp. 1428–1431, September 2012.
- [5] K. Zheng, F. Hu, W. Wang, W. Xiang, and M. Dohler, "Radio resource allocation in LTE-advanced cellular networks with M2M communications," *IEEE Commun. Mag.*, vol. 50, no. 7, pp. 184–192, July 2012.
- [6] A. G. Gotsis, A. S. Lioumpas, and A. Alexiou, "Analytical modelling and performance evaluation of realistic time-controlled M2M scheduling over LTE cellular networks," *Transactions on Emerging Telecommunications Technologies*, vol. 24, no. 4, pp. 378–388, 2013.
- [7] M. Hasan, E. Hossain, and D. Niyato, "Random access for machine-to-machine communication in LTE-advanced networks: issues and approaches," *IEEE Commun. Mag.*, vol. 51, no. 6, pp. 86–93, 2013.
- [8] A. Lo, Y. W. Law, M. Jacobsson, and M. Kucharzak, "Enhanced LTE-advanced random-access mechanism for massive machine-to-machine (M2M) communications," in *27th Wireless World Research Forum (WWRF27)*, 2011.
- [9] S. Choi, W. Lee, D. Kim, K.-J. Park, S. Choi, and K.-Y. Han, "Automatic configuration of random access channel parameters in LTE systems," in *2011 IFIP Wireless Days (WD)*.
- [10] C.-Y. Tu, C.-Y. Ho, and C.-Y. Huang, "Energy-efficient algorithms and evaluations for massive access management in cellular based machine to machine communications," in *IEEE Vehicular Technology Conference (VTC Fall)*, 2011.
- [11] J.-P. Cheng, C. han Lee, and T.-M. Lin, "Prioritized random access with dynamic access barring for RAN overload in 3GPP LTE-A networks," in *Proc. IEEE GLOBECOM Workshops (GC Wkshps)*, 2011, pp. 368–372.
- [12] S.-Y. Lien, T.-H. Liao, C.-Y. Kao, and K.-C. Chen, "Cooperative access class barring for machine-to-machine communications," *IEEE Trans. Wireless Commun.*, vol. 11, no. 1, pp. 27–32, 2012.
- [13] H. S. Dhillon, H. C. Huang, H. Viswanathan, and R. A. Valenzuela, "Fundamentals of throughput maximization with random arrivals for M2M communications," *IEEE Trans. Commun.*, vol. 62, no. 11, pp. 4094–4109, Nov. 2014.
- [14] —, "Power-efficient system design for cellular-based machine-to-machine communications," *IEEE Trans. Wireless Commun.*, vol. 12, no. 11, pp. 5740–5753, Nov. 2013.
- [15] M. Shirvanimoghaddam, Y. Li, and B. Vucetic, "Near-capacity adaptive analog fountain codes for wireless channels," *IEEE Commun. Lett.*, Apr. 2013.
- [16] K. Wu, Z. Zhang, and S. Chen, "Rateless multiple access over erasure channel," in *IEEE 71st Vehicular Technology Conference (VTC 2010-Spring)*, 2010, pp. 1–6.
- [17] A. Khisti, B. Hern, and K. Narayanan, "On modulo-sum computation over an erasure multiple access channel," in *Proc. IEEE Intl. Symp. Inform. Theory (ISIT)*, 2012, pp. 3013–3017.
- [18] A. Tehrani, A. Dimakis, and M. Neely, "Sigsag: Iterative detection through soft message-passing," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 8, pp. 1512–1523, 2011.
- [19] G. Liva, E. Paolini, M. Lentmaier, and M. Chiani, "Spatially-coupled random access on graphs," in *Proc. IEEE Intl. Symp. Inform. Theory (ISIT)*, 2012, pp. 478–482.
- [20] D. Truhachev, "Achieving Gaussian multiple access channel capacity with spatially coupled sparse graph multi-user modulation," in *Proc. Information Theory and Applications Workshop (ITA)*, 2012, pp. 331–337.
- [21] M. Shirvanimoghaddam, Y. Li, and B. Vucetic, "Multiple access analog fountain codes," in *Proc. IEEE Intl. Symp. Inform. Theory (ISIT)*, Jun. 2014.
- [22] A. Laya, L. Alonso, and J. Alonso-Zarate, "Is the random access channel of LTE and LTE-A suitable for M2M communications? a survey of alternatives," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 1, pp. 4–16, First 2014.
- [23] H. Moon, "Optimum power allocation for preamble detection with channel-adaptive random access," *IEEE Trans. Wireless Commun.*, vol. 12, no. 11, pp. 5424–5433, Nov. 2013.
- [24] R. Louie, W. Hardjawana, Y. Li, and B. Vucetic, "Distributed multiple-access for wireless communications: Compressed sensing with multiple antennas," in *IEEE Global Communications Conference (GLOBECOM)*, Dec. 2012, pp. 3622–3627.
- [25] D. Niyato, P. Wang, and D. I. Kim, "Performance modeling and analysis of heterogeneous machine type communications," *IEEE Trans. Wireless Commun.*, vol. 13, no. 5, pp. 2836–2849, May 2014.
- [26] "Machine-to-machine communications (M2M); applicability of M2M architecture to smart grid networks; impact of smart grids on M2M platform," ETSI, Technical Report DTR/M2M-00011, Sep. 2012.
- [27] M. Shirvanimoghaddam, Y. Li, and B. Vucetic, "Adaptive analog fountain for wireless channels," in *Proceedings. IEEE Wireless Communications and Networking Conference (WCNC)*, Apr. 2013.
- [28] Y. Kabashima, "A CDMA multiuser detection algorithm on the basis of belief propagation," *Journal of Physics A: Mathematical and General*, vol. 36, no. 43, p. 11111, 2003.
- [29] M. Shirvanimoghaddam, Y. Li, S. Tian, and B. Vucetic, "Distributed rap-

- tor coding for erasure channels: Partially and fully coded cooperation,” *IEEE Trans. Commun.*, vol. 26, no. 3, pp. 421–431, 2013.
- [30] N. Rahnavard, B. N. Vellambi, and F. Fekri, “Rateless codes with unequal error protection property,” *IEEE Transactions on Information Theory*, vol. 53, pp. 1521–1532, Apr. 2007.
  - [31] M. Luby, “LT codes,” in *Proc. 43rd Annual IEEE Symp. Foundations Comput. Science*, Nov. 2002, pp. 271 – 280.
  - [32] K. Zheng, S. Ou, J. Alonso-Zarate, M. Dohler, F. Liu, and H. Zhu, “Challenges of massive access in highly dense lte-advanced networks with machine-to-machine communications,” *IEEE Wireless Commun. Mag.*, vol. 21, no. 3, pp. 12–18, Jun. 2014.
  - [33] D. Guo and C.-C. Wang, “Multiuser detection of sparsely spread CDMA,” *IEEE J. Sel. Areas Commun.*, vol. 26, no. 3, pp. 421–431, 2008.